ELSEVIER

Discrete Applied Mathematics 111 (2001) 263–281

DISCRETE APPLIED MATHEMATICS

On the minimum average distance of binary codes: linear programming approach ☆

Fang-Wei Fu[a,*], Victor K. Wei[b], Raymond W. Yeung[b]

[a] *Department of Mathematics, Nankai University, Tianjin 300071, China*

[b] *Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

Received 26 January 1999; revised 23 May 2000; accepted 5 June 2000

Abstract

Ahlswede and Katona posed the following average distance problem: For every n and $1 \leqslant M \leqslant 2^n$, determine the minimum average Hamming distance $\beta(n, M)$ of binary codes with length n and size M. In this paper, improved lower bounds for $\beta(n, M)$ are found with the help of linear programming. As a corollary, $\beta(n, 2^{n-2} \pm 1)$, $\beta(n, 2^{n-1} + 2^{n-2} \pm 1)$, $\beta(n, 2^{n-2})$, $\beta(n, 2^{n-1} + 2^{n-2})$, $\beta(n, 2^{n-1} \pm 2)$ and $\beta(n, 2^n - 2)$ are determined. Furthermore, an upper bound for $\beta(n, M)$ is obtained by constructing a binary code with length n and size M. This upper bound is tight for some cases, but not in general.

Keywords: Binary codes; Average Hamming distance; Distance enumerator; MacWilliams–Delsarte identity; Linear programming problem

1. Introduction

Let $V_n = \{0, 1\}^n$ be the n-dimensional vector space over the binary field $\{0, 1\}$. The Hamming distance between two vectors a and b is the number of components where they differ, and is denoted by $d_H(a, b)$. The Hamming weight of a vector x is the number of nonzero components, and is denoted by $w_H(x)$. For $x = (x_1, \ldots, x_n) \in V_n$ and $y = (y_1, \ldots, y_n) \in V_n$, the scalar product of x and y is defined by

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

☆ This research work is supported in part by the National Natural Science Foundation of China under the Grant 69802008, the Foundation for University Key Teacher by the Education Ministry of China, and the Research Grant Council of Hong Kong under Earmarked Grant CUHK 4424/99E and 332/96E.

* Correspondence address: Department of Mathematics, Nankai University, Tianjin 300071, China.

E-mail addresses: fwfu@nankai.edu.cn (F.-W. Fu), kwwei@ie.cuhk.edu.hk (V.K. Wei), whyeung@ie.cuhk.edu.hk (R.W. Yeung).

0166-218X/01/$ - see front matter
PII: S0166-218X(00)00284-5

A subset C of V_n with size M is called a binary (n,M) code. The average Hamming distance of C is defined by

$$\bar{d}(C)=\frac{1}{M^2}\sum_{a\in C}\sum_{b\in C}d_{\mathrm{H}}(a,b). \tag{1.1}$$

The minimum average Hamming distance of a binary (n,M) code is defined by

$$\beta(n,M)=\min\{\bar{d}(C)\,|\,C \text{ is a binary } (n,M) \text{ code}\}. \tag{1.2}$$

Ahlswede and Katona [2] posed the following problem on the extremal combinatorics of Hamming space: For every $1\leqslant M\leqslant 2^n$, determine the exact value of $\beta(n,M)$. Ahlswede and Althöfer [1] observed that this problem also occurs in the construction of good codes for write-efficient memories, introduced by Ahlswede and Zhang [3] as a model for storing and updating information on a rewritable medium with cost constraints. One problem that arises is to find sets $C\subseteq V_n$ of a given cardinality which minimize the average inner cost

$$\frac{1}{|C|^2}\sum_{a\in C}\sum_{b\in C}d_{\mathrm{H}}(a,b).$$

Kündgen [9] also observed that this problem is equivalent to a covering problem in graph theory.

A referee pointed out that for every subset $C\subseteq V_n$, there is a relationship between $\bar{d}(C)$ and $\bar{d}(V_n\backslash C)$:

Lemma 1. *For every subset $C\subseteq V_n$, we have*

$$\bar{d}(V_n\backslash C)=\frac{n}{2}-\frac{|C|^2}{(2^n-|C|)^2}\left[\frac{n}{2}-\bar{d}(C)\right]. \tag{1.3}$$

As pointed out by the referee, Lemma 1 implies that

Lemma 2. *For a subset $C\subseteq V_n$, if $\bar{d}(C)=\beta(n,|C|)$, then $\bar{d}(V_n\backslash C)=\beta(n,2^n-|C|)$. Furthermore, for $1\leqslant M\leqslant 2^n$,*

$$\beta(n,2^n-M)=\frac{n}{2}-\frac{M^2}{(2^n-M)^2}\left[\frac{n}{2}-\beta(n,M)\right]. \tag{1.4}$$

Lemma 2 implies that we only need to determine $\beta(n,M)$ for $1\leqslant M\leqslant 2^{n-1}$.

For completeness, here we present a proof for Lemma 1 based on the definitions. In Section 2, we will give another proof for Lemma 1 based on the properties of the distance distribution of codes.

Proof. Let A be a binary $|C|\times n$ matrix, whose rows consist of all vectors in C. For $i=1,2,\ldots,n$, suppose that the number of 1's in the ith column of A is m_i. Then

$$I_1=\sum_{a\in C}\sum_{b\in C}d_{\mathrm{H}}(a,b)=2\sum_{i=1}^{n}m_i(|C|-m_i). \tag{1.5}$$

Let $\bar{A}$ be a binary $(2^n-|C|)\times n$ matrix, whose rows consist of all vectors in $V_n\backslash C$. We know that for $i=1,2,\ldots,n$, the number of 1's in the ith column of $\bar{A}$ is $(2^{n-1}-m_i)$. Then

$$\begin{aligned} I_2 &= \sum_{a\in V_n\backslash C}\sum_{b\in V_n\backslash C} d_{\mathrm{H}}(a,b) \\ &= 2\sum_{i=1}^{n}(2^{n-1}-m_i)(2^{n-1}+m_i-|C|) \\ &= 2n(2^{n-1})^2 - n2^n|C| + 2\sum_{i=1}^{n} m_i(|C|-m_i) \\ &= (n/2)[(2^n-|C|)^2-|C|^2]+I_1. \end{aligned} \tag{1.6}$$

Since $\bar{d}(V_n\backslash C)=I_2/(2^n-|C|)^2$ and $\bar{d}(C)=I_1/|C|^2$, we have

$$\begin{aligned} \bar{d}(V_n\backslash C) &= \frac{n}{2} - \frac{n}{2}\frac{|C|^2}{(2^n-|C|)^2} + \frac{|C|^2}{(2^n-|C|)^2}\bar{d}(C) \\ &= \frac{n}{2} - \frac{|C|^2}{(2^n-|C|)^2}\left[\frac{n}{2}-\bar{d}(C)\right]. \end{aligned}$$

This completes the proof of Lemma 1. □

In the process of making efforts to solve the open problem posed by Ahlswede et al., Althöfer and Sillke [4] proved that

Theorem 1 (Althöfer and Sillke [4]). *Let C be a binary (n,M) code, then*

$$\bar{d}(C)\geqslant\frac{n+1}{2}-\frac{2^{n-1}}{M}, \tag{1.7}$$

where equality is possible only for $M=2^n$ and for $M=2^{n-1}$ with C being a subcube.

Theorem 1 shows that

$$\beta(n,2^n)=\frac{n}{2},\quad \beta(n,2^{n-1})=\frac{n-1}{2}. \tag{1.8}$$

Xia and Fu [11] improved Theorem 1 for odd M as follows.

Theorem 2 (Xia and Fu [1]). *Let C be a binary (n,M) code. If M is odd, then*

$$\bar{d}(C)\geqslant\frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{2^n-n-1}{2M^2}, \tag{1.9}$$

where equality holds for $M=2^n-1$ with C being a set obtained by removing one point from V_n, and for $M=2^{n-1}\pm 1$ with C being a set obtained by adding or removing one point from a subcube.

Theorem 2 shows that

$$\beta(n, 2^n - 1) = \frac{n}{2} - \frac{n}{2(2^n - 1)^2}, \tag{1.10}$$

$$\beta(n, 2^{n-1} - 1) = \frac{n-1}{2} - \frac{n-1}{2(2^{n-1} - 1)^2}, \tag{1.11}$$

$$\beta(n, 2^{n-1} + 1) = \frac{n-1}{2} + \frac{2^{n+1} - n + 1}{2(2^{n-1} + 1)^2}. \tag{1.12}$$

Ahlswede and Althöfer [1] studied the asymptotic behaviour of $\beta(n, M)$. In general, the expectation and variance of two independent identical distributed random vectors over $GF(2)$ and $GF(q)$ are studied by Althöfer and Sillke [4], Fu and Shen [8], and Fu et al. [7].

In this paper, stimulated by Delsarte's linear programming bound [6] for codes, we find several improved lower bounds for $\beta(n, M)$ with the help of linear programming. As a corollary, $\beta(n, 2^{n-2} \pm 1), \beta(n, 2^{n-1} + 2^{n-2} \pm 1), \beta(n, 2^{n-2}), \beta(n, 2^{n-1} + 2^{n-2}), \beta(n, 2^{n-1} \pm 2)$ and $\beta(n, 2^n - 2)$ are determined. Furthermore, an upper bound for $\beta(n, M)$ is obtained by constructing a binary (n, M) code. This upper bound is tight for some cases, but not in general.

This paper is organized as follows. In Section 2, we first review some basic properties of Krawtchouk polynomials, distance distribution and dual distance distribution of codes. Then we mention some basic results regarding linear programming problem. In Section 3, we present a lower bound for $\beta(n, M)$ with $M \equiv 2 \pmod 4$. This enable us to determine $\beta(n, 2^n - 2)$ and $\beta(n, 2^{n-1} \pm 2)$. In Section 4, by using the linear programming technique, we present the linear programming bounds for $\beta(n, M)$. Several explicit lower bounds for $\beta(n, M)$ are derived by using these linear programming bounds and Lemma 1. These lower bounds improve the previously known lower bounds. As a corollary, $\beta(n, 2^{n-2})$, $\beta(n, 2^{n-1} + 2^{n-2})$, $\beta(n, 2^{n-2} \pm 1)$ and $\beta(n, 2^{n-1} + 2^{n-2} \pm 1)$ are determined. In Section 5, we present an upper bound for $\beta(n, M)$. This upper bound is tight for some cases, but not in general.

2. Preliminaries

In this section, we review some basic properties regarding the Krawtchouk polynomials, the distance distribution of codes and the linear programming problems. For details we refer the readers to [10]. We will use these properties to establish our results. We also present a new proof for Lemma 1 based on the properties of the distance distribution of codes.

2.1. Krawtchouk polynomials

Here we mention several properties of Krawtchouk polynomials that will be used in the rest of the paper. For details we refer to [10, Section 1.2].

Let $\mathfrak{R} = (-\infty, +\infty)$. For $k = 0, 1, 2, \ldots$, the Krawtchouk polynomial $K_k(z; n)$ is defined by

$$K_k(z; n) = \sum_{j=0}^{k} (-1)^j \binom{z}{j} \binom{n-z}{k-j}, \quad \text{where } z \in \mathfrak{R}. \tag{2.1}$$

If the parameter n is clear from the context, we simply write $K_k(z)$ instead of $K_k(z; n)$. In this paper we only need to consider the case where z is an integer and $0 \leqslant z \leqslant n$.

The Krawtchouk polynomials satisfy the following relations:

$$K_k(0) = \binom{n}{k}, \tag{2.2}$$

$$K_k(i) = (-1)^i K_{n-k}(i), \quad i = 0, 1, \ldots, n, \tag{2.3}$$

$$K_k(n-i) = (-1)^k K_k(i), \tag{2.4}$$

$$\sum_{i=0}^{n} \binom{n}{i} K_k(i) K_l(i) = \delta_{kl} \binom{n}{k} 2^n, \tag{2.5}$$

where $\delta_{kl} = 0$, $k \neq l$ and $\delta_{kl} = 1$, $k = l$,

$$K_n(i) = (-1)^i, \quad i = 0, 1, \ldots, n, \tag{2.6}$$

$$K_k(1) = \left[1 - \frac{2k}{n}\right] \binom{n}{k}, \tag{2.7}$$

$$\sum_{k=0}^{n} K_k(l) = 0, \quad \text{where } 1 \leqslant l \leqslant n. \tag{2.8}$$

For any nonnegative integers r, s,

$$\binom{n}{i} K_s(i) = \binom{n}{s} K_i(s), \tag{2.9}$$

$$\sum_{i=0}^{n} K_r(i) K_i(s) = 2^n \delta_{r,s}. \tag{2.10}$$

The Krawtchouk polynomials of degree up to three are:

$$K_0(z) = 1, \tag{2.11}$$

$$K_1(z) = n - 2z, \tag{2.12}$$

$$K_2(z) = \binom{n}{2} - 2nz + 2z^2, \tag{2.13}$$

$$K_3(z) = \binom{n}{3} - \left(n^2 - n + \frac{2}{3}\right) z + 2nz^2 - \frac{4}{3}z^3. \tag{2.14}$$

2.2. Distance distributions

Let C be a binary (n, M) code. The distance distribution of C is defined by

$$A_i = \frac{1}{M}|\{(a,b) \mid a,b \in C, d_H(a,b) = i\}|, \quad i = 0,1,\ldots,n. \tag{2.15}$$

The dual distance distribution of C is defined by

$$B_i = \frac{1}{M^2} \sum_{\substack{u \in V_n \\ w_H(u)=i}} \left[\sum_{c \in C} (-1)^{\langle u,c\rangle}\right]^2, \quad i = 0,1,\ldots,n. \tag{2.16}$$

The distance enumerator of C is defined as $f(s) = \sum_{i=0}^{n} A_i s^i$, and the dual distance enumerator of C is defined as $g(s) = \sum_{i=0}^{n} B_i s^i$. The MacWilliams–Delsarte identity (see [10]) gives the relationship between $f(s)$ and $g(s)$ as follows:

$$g(s) = \frac{1}{M}(1+s)^n f\left(\frac{1-s}{1+s}\right), \tag{2.17}$$

$$f(s) = \frac{M}{2^n}(1+s)^n g\left(\frac{1-s}{1+s}\right). \tag{2.18}$$

It is easy to see from the MacWilliams–Delsarte identity (or the Pless identity for the moments of distance distribution) that

Property 1. (Xia and Fu [11])

$$\bar{d}(C) = \frac{n}{2} - \frac{B_1}{2}. \tag{2.19}$$

From Property 1, we can give a new proof for Lemma 1 as follows.

A new Proof of Lemma 1. For a nonempty subset $C \subseteq V_n$, denote $\bar{C} = V_n \setminus C$. The distance distributions of C and $\bar{C}$ are given by $\{B_i\}_{i=0}^n$ and $\{\bar{B}_i\}_{i=0}^n$, respectively. We know from [10] that for every nonzero vector $u \in V_n$,

$$\sum_{a \in V_n} (-1)^{\langle u,a\rangle} = 0.$$

This implies that

$$\sum_{a \in C} (-1)^{\langle u,a\rangle} = -\sum_{a \in \bar{C}} (-1)^{\langle u,a\rangle}.$$

It follows from (2.16) that for $i = 1, 2, \ldots, n$,

$$\begin{aligned} |C|^2 B_i &= \sum_{\substack{u \in V_n \\ w_H(u)=i}} \left[\sum_{a \in C} (-1)^{\langle u,a \rangle} \right]^2 \\ &= \sum_{\substack{u \in V_n \\ w_H(u)=i}} \left[\sum_{a \in \bar{C}} (-1)^{\langle u,a \rangle} \right]^2 \\ &= (2^n - |C|)^2 \bar{B}_i. \end{aligned}$$

From Property 1, we obtain that

$$\begin{aligned} \bar{d}(V_n \setminus C) &= \frac{n}{2} - \frac{\bar{B}_1}{2} \\ &= \frac{n}{2} - \frac{|C|^2}{(2^n - |C|)^2} \frac{B_1}{2} \\ &= \frac{n}{2} - \frac{|C|^2}{(2^n - |C|)^2} \left[\frac{n}{2} - \bar{d}(C) \right]. \end{aligned}$$

This completes the proof of Lemma 1. □

We will use the following properties of distance distribution.

Property 2 . (MacWilliams and Sloane [10])

$$B_0 = 1, \quad B_i \geqslant 0, \quad i = 1, 2, \ldots, n, \tag{2.20}$$

$$\sum_{i=0}^{n} B_i = \frac{2^n}{M}. \tag{2.21}$$

Property 3 . (Best et al. [5])

If $1 \leqslant M \leqslant 2^n$ *and* M *is odd, then*

$$B_i \geqslant \frac{1}{M^2} \binom{n}{i}, \quad i = 0, 1, \ldots, n. \tag{2.22}$$

Property 4 . (Best et al. [5])

If $1 \leqslant M \leqslant 2^n$ *and* $M \equiv 2 \pmod 4$, *then there exits* $l \in \{0, 1, \ldots, n\}$, *such that*

$$B_i \geqslant \frac{2}{M^2} \left[\binom{n}{i} + K_i(l) \right], \quad k = 0, 1, \ldots, n, \tag{2.23}$$

where $K_i(z)$ *is the Krawtchouk polynomial defined by* (2.1).

The MacWilliams–Delsarte identity also gives the relationship between the distance distribution and the dual distance distribution.

Property 5 . (Best et al. [5])

$$B_k = \frac{1}{M}\sum_{i=0}^{n} K_k(i)A_i, \quad k = 0,1,\ldots,n, \tag{2.24}$$

$$A_k = \frac{M}{2^n}\sum_{i=0}^{n} K_k(i)B_i, \quad k = 0,1,\ldots,n. \tag{2.25}$$

2.3. Linear programming problem

Here we review some basic concepts, notations and properties of linear programming. We will use linear programming to obtain lower bounds for $\beta(n,M)$. Linear programming is a technique for maximizing (or minimizing) a linear form, called the objective function, subject to certain linear constraints. For details we refer to [10, Section 17.4].

Problem I . (The primal linear programming problem)
Choose the real variables $x_1, x_2, \ldots, x_s$ so as to maximize the objective function

$$\sum_{j=1}^{s} c_j x_j \tag{2.26}$$

subject to the inequalities

$$x_j \geqslant 0, \quad j = 1,2,\ldots,s, \tag{2.27}$$

$$\sum_{j=1}^{s} a_{ij}x_j \geqslant -b_i, \quad i = 1,2,\ldots,n. \tag{2.28}$$

Problem II . (The dual linear programming problem)
Choose the real variables $u_1, u_2, \ldots, u_n$ so as to minimize

$$\sum_{i=1}^{n} u_i b_i \tag{2.29}$$

subject to the inequalities

$$u_i \geqslant 0, \quad i = 1,2,\ldots,n, \tag{2.30}$$

$$\sum_{i=1}^{n} u_i a_{ij} \leqslant -c_j, \quad j = 1,2,\ldots,s. \tag{2.31}$$

Let $A=(a_{ij})_{n\times s}$, $b=(b_1,\ldots,b_n)$, $c=(c_1,\ldots,c_s)$, $x=(x_1,\ldots,x_s)$, $u=(u_1,\ldots,u_n)$. These problems can be restated in matrix notation as follows:

(I′) Maximize cx^{T} subject to $x\geqslant 0$, $Ax^{\mathrm{T}}\geqslant -b^{\mathrm{T}}$,

(II′) Minimize ub^{T} subject to $u\geqslant 0$, $uA\leqslant -c$.

A vector x (resp. u) is called a *feasible solution* to Problem I (resp. Problem II), if it satisfies the inequalities of Problem I (resp. Problem II), and an *optimal solution* if it also maximizes cx^{T} (resp. minimizes ub^{T}).

Property 6. *If x and u are feasible solutions to Problem* I *and* II, *respectively*, *then* $cx^{\mathrm{T}}\leqslant ub^{\mathrm{T}}$.

Property 7. *If x and u are feasible solutions to Problem* I *and* II, *respectively*, *then x and u are both optimal iff* $cx^{\mathrm{T}}=ub^{\mathrm{T}}$.

3. A lower bound for $\beta(n, M)$ with $M\equiv 2(\mathrm{mod}4)$

Theorem 3. *Let C be a binary (n,M) code. If* $M\equiv 2\,(\mathrm{mod}\,4)$, *then*

$$\bar{d}(C)\geqslant \frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{2^n-2n}{M^2}\triangleq \Delta_1(n,M). \tag{3.1}$$

Proof. Let $\{B_k\}_{k=0}^{n}$ be the dual distance distribution of code C, then by Properties 1 and 2, we have

$$\begin{aligned}\bar{d}(C)&=\frac{n}{2}-\frac{B_1}{2}\\&=\frac{n}{2}-\frac{1}{2}\left(\frac{2^n}{M}-1-\sum_{k=2}^{n}B_k\right)\\&=\frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{1}{2}\sum_{k=2}^{n}B_k.\end{aligned} \tag{3.2}$$

From Property 4, we know that there exists $l\in\{0,1,\ldots,n\}$ such that

$$\begin{aligned}\sum_{k=2}^{n}B_k&\geqslant\frac{2}{M^2}\sum_{k=2}^{n}\left[\binom{n}{k}+K_k(l)\right]\\&=\frac{2}{M^2}\left[2^n-1-n+\sum_{k=2}^{n}K_k(l)\right].\end{aligned} \tag{3.3}$$

If $l=0$, then by (2.2),

$$\sum_{k=2}^{n}K_k(l)=\sum_{k=2}^{n}\binom{n}{k}=2^n-1-n. \tag{3.4}$$

If $l \geqslant 1$, then by (2.8), (2.11) and (2.12),

$$\sum_{k=2}^{n} K_k(l) = \sum_{k=0}^{n} K_k(l) - K_0(l) - K_1(l)$$
$$= -1 - (n-2l) = -1 - n + 2l \geqslant -(n-1). \tag{3.5}$$

Since $2^n - 1 - n \geqslant -(n-1)$, we have

$$\sum_{k=2}^{n} K_k(l) \geqslant -(n-1). \tag{3.6}$$

By (3.2), (3.3) and (3.6), we have

$$\bar{d}(C) \geqslant \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{1}{M^2}[2^n - 1 - n - (n-1)]$$
$$= \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{2^n - 2n}{M^2}. \qquad \square$$

Note that Theorem 3 is only meaningful for $\Delta_1(n, M) > 0$. The following corollary shows that Theorem 3 is tight for some cases.

Corollary 1.

$$\beta(n, 2^n - 2) = \frac{n}{2} - \frac{2n-2}{(2^n-2)^2}, \tag{3.7}$$

$$\beta(n, 2^{n-1} - 2) = \frac{n-1}{2} - \frac{2n-4}{(2^{n-1}-2)^2}, \tag{3.8}$$

$$\beta(n, 2^{n-1} + 2) = \frac{n-1}{2} + \frac{2^{n+1} - 2n + 4}{(2^{n-1}+2)^2}. \tag{3.9}$$

Proof. Let $\mathbf{0}$ be the zero vector with length n. Let e_i be the binary vector with length n in which only the ith coordinate is 1. Let

$$C_1 = V_n \setminus \{\mathbf{0}, e_1\},$$
$$C_2 = V_{n-1} \times \{0\} \setminus \{\mathbf{0}, e_1\},$$
$$C_3 = V_{n-1} \times \{0\} \cup \{e_n, e_1 + e_n\}.$$

By direct calculation, it is not hard to obtain that

$$\bar{d}(C_1) = \frac{n}{2} - \frac{2n-2}{(2^n-2)^2},$$
$$\bar{d}(C_2) = \frac{n-1}{2} - \frac{2n-4}{(2^{n-1}-2)^2},$$
$$\bar{d}(C_3) = \frac{n-1}{2} + \frac{2^{n+1} - 2n + 4}{(2^{n-1}+2)^2}.$$

Since $|C_1| = 2^n - 2$, $|C_2| = 2^{n-1} - 2$ and $|C_3| = 2^{n-1} + 2$, we have

$$\beta(n, 2^n - 2) \leqslant \frac{n}{2} - \frac{2n-2}{(2^n-2)^2},$$
$$\beta(n, 2^{n-1} - 2) \leqslant \frac{n-1}{2} - \frac{2n-4}{(2^{n-1}-2)^2},$$
$$\beta(n, 2^{n-1} + 2) \leqslant \frac{n-1}{2} + \frac{2^{n+1}-2n+4}{(2^{n-1}+2)^2}.$$

It is easy to see from the lower bound of Theorem 3 that

$$\beta(n, 2^n - 2) \geqslant \frac{n}{2} - \frac{2n-2}{(2^n-2)^2},$$
$$\beta(n, 2^{n-1} - 2) \geqslant \frac{n-1}{2} - \frac{2n-4}{(2^{n-1}-2)^2},$$
$$\beta(n, 2^{n-1} + 2) \geqslant \frac{n-1}{2} + \frac{2^{n+1}-2n+4}{(2^{n-1}+2)^2}.$$

Corollary 1 follows by combining these assertions. □

Remark. By Lemma 2, we can also determine $\beta(n, 2^n - 2)$ and $\beta(n, 2^{n-1} + 2)$ from $\beta(n, 2)$ and $\beta(n, 2^{n-1} - 2)$, respectively. Here we have shown that the lower bound of Theorem 3 is tight for $M = 2^n - 2, 2^{n-1} \pm 2$.

4. Linear programming (LP) bounds

In this section, we present several new lower bounds for $\beta(n, M)$ by using the linear programming technique. Furthermore, we show that these lower bounds are tight for some cases.

4.1. LP bound for $\beta(n, M)$ with $M \leqslant 2^{n-1}$

Let C be a binary (n, M) code. From (3.2), we know that

$$\bar{d}(C) = \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{1}{2}\sum_{k=2}^{n} B_k.$$

From Properties 2 and 5, we know that

$$B_0 = 1, \quad B_i \geqslant 0, \quad i = 1, 2, \ldots, n, \tag{4.1}$$

$$\sum_{i=0}^{n} K_k(i) B_i \geqslant 0, \quad k = 1, 2, \ldots, n. \tag{4.2}$$

Substituting

$$B_1 = \frac{2^n}{M} - 1 - B_2 - B_3 - \cdots - B_n,$$

$$K_k(0) = \binom{n}{k}, \quad K_k(1) = \left[1 - \frac{2k}{n}\right]\binom{n}{k}$$

into (4.2), we obtain that

$$\sum_{i=2}^{n}\left[\binom{n}{k}\left(1-\frac{2k}{n}\right) - K_k(i)\right]B_i \leqslant \binom{n}{k}\left[1+\left(1-\frac{2k}{n}\right)\left(\frac{2^n}{M}-1\right)\right], \quad k=1,2,\ldots,n. \tag{4.3}$$

Consider the following linear programming (LP) problem:

($\mathscr{A}$) Choose the real variables $u_2,\ldots,u_n$ so as to

$$\Lambda_1(n,M) = \text{minimize}\sum_{i=2}^{n} u_i \tag{4.4}$$

subject to the inequalities

$$u_i \geqslant 0, \quad i=2,3,\ldots,n, \tag{4.5}$$

$$\sum_{i=2}^{n}\left[\binom{n}{k}\left(1-\frac{2k}{n}\right) - K_k(i)\right]u_i \leqslant \binom{n}{k}\left[1+\left(1-\frac{2k}{n}\right)\left(\frac{2^n}{M}-1\right)\right], \quad k=1,2,\ldots,n. \tag{4.6}$$

Note that if $M \geqslant 2^{n-1}$, then

$$1+\left(1-\frac{2k}{n}\right)\left(\frac{2^n}{M}-1\right) \geqslant 0, \quad k=1,2,\ldots,n. \tag{4.7}$$

This implies that $\Lambda_1(n,M)=0$ by choosing the optimal solution $u_2=u_3=\cdots=u_n=0$. Hence, here we only consider the case $M \leqslant 2^{n-1}$. By (3.2) we obtain a LP bound for $\beta(n,M)$ with $M \leqslant 2^{n-1}$ as follows:

$$\bar{d}(C) \geqslant \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{1}{2}\Lambda_1(n,M). \tag{4.8}$$

The dual problem of ($\mathscr{A}$) is given as follows (see Section 2.3, Properties 6 and 7):

($\mathscr{A}'$) Choose the real variables $x_1,x_2,\ldots,x_n$ so as to

$$\Lambda_1(n,M) = \text{maximize}\sum_{k=1}^{n}\binom{n}{k}\left[\left(\frac{2k}{n}-1\right)\left(\frac{2^n}{M}-1\right)-1\right]x_k \tag{4.9}$$

subject to the inequalities

$$x_k \geqslant 0, \quad k = 1, 2, \ldots, n, \tag{4.10}$$

$$\sum_{k=1}^{n} \left[\binom{n}{k} \left(1 - \frac{2k}{n} \right) - K_k(i) \right] x_k \geqslant -1, \quad i = 2, 3, \ldots, n. \tag{4.11}$$

It is not difficult to see that $x_1 = x_2 = \cdots = x_{n-1} = 0$, $x_n = \frac{1}{2}$ is a feasible solution of the LP problem $(\mathscr{A}')$. Here we only need to check that

$$[-1 - (-1)^i]\frac{1}{2} \geqslant -1.$$

Hence

$$\Lambda_1(n, M) \geqslant \frac{1}{2} \left[\frac{2^n}{M} - 1 - 1 \right] = \frac{2^{n-1}}{M} - 1.$$

Therefore it follows from (4.8) that

Theorem 4. *Let C be a binary (n, M) code, then for $M \leqslant 2^{n-1}$,*

$$\bar{d}(C) \geqslant \frac{n}{2} - \frac{2^{n-2}}{M}. \tag{4.12}$$

Note that the lower bound in Theorem 4 is only meaningful for $2^{n-1}/n \leqslant M \leqslant 2^{n-1}$. It follows from (1.8) that the lower bound in Theorem 4 is tight for $M = 2^{n-1}$. The following corollary shows that the lower bound in Theorem 4 is also tight for $M = 2^{n-2}$.

Corollary 2.

$$\beta(n, 2^{n-2}) = \frac{n-2}{2}. \tag{4.13}$$

Proof. Theorem 4 implies that $\beta(n, 2^{n-2}) \geqslant (n-2)/2$. Let $C = V_{n-2} \times \{00\}$, then it is easy to see that $|C| = 2^{n-2}$ and $\bar{d}(C) = (n-2)/2$. Therefore $\beta(n, 2^{n-2}) \leqslant (n-2)/2$ and the corollary follows. □

It follows from Lemma 2 and Corollary 2 that

Corollary 3.

$$\beta(n, 2^{n-1} + 2^{n-2}) = \frac{n}{2} - \frac{1}{9}. \tag{4.14}$$

From Lemma 1 and Theorem 4, we obtain that

Corollary 4. *Let C be a binary (n, M) code, then for $M \geqslant 2^{n-1}$,*

$$\bar{d}(C) \geqslant \frac{n}{2} - \frac{2^{n-2}(2^n - M)}{M^2}. \tag{4.15}$$

From (1.8) and (4.14), we know that the lower bound in Corollary 4 is tight for $M=2^{n-1}, 2^{n-1}+2^{n-2}$. Comparing the lower bounds in Theorem 4 and Corollary 4 with the lower bound of Althöfer and Sillke (see Theorem 1), we see that both the lower bounds in Theorem 4 and Corollary 4 are better than the lower bound of Althöfer and Sillke.

4.2. LP bound for $\beta(n,M)$ with M odd and $M \leqslant 2^{n-1}-1$

Let C be a binary (n,M) code. If M is odd, then from Property 3 we know that

$$B_i \geqslant \frac{1}{M^2}\binom{n}{i}, \quad i=0,1,\ldots,n.$$

Let $B'_i = B_i - (1/M^2)\binom{n}{i}$, $i=0,1,\ldots,n$. It follows from (3.2) that

$$\begin{aligned}\bar{d}(C) &= \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{1}{2M^2}\sum_{i=2}^{n}\binom{n}{i} + \frac{1}{2}\sum_{i=2}^{n} B'_i \\ &= \frac{n+1}{2} - \frac{2^{n-1}}{M} + \frac{2^n-n-1}{2M^2} + \frac{1}{2}\sum_{i=2}^{n} B'_i. \end{aligned} \tag{4.16}$$

Substituting $B_i = B'_i + (1/M^2)\binom{n}{i}$ into (4.3), and using the properties of Krawtchouk polynomials, we obtain that

$$\begin{aligned}&\sum_{i=2}^{n}\left[\binom{n}{k}\left(1-\frac{2k}{n}\right) - K_k(i)\right] B'_i \\ &\quad \leqslant \binom{n}{k}\left[1 - \frac{1}{M^2} + (1-\frac{2k}{n})(\frac{2^n}{M} - 1 - \frac{2^n-1}{M^2})\right], \quad k=1,2,\ldots,n. \end{aligned} \tag{4.17}$$

Consider the following linear programming problem:

($\mathscr{B}$) Choose the real variables $u_2,\ldots,u_n$ so as to

$$\Lambda_2(n,M) = \text{minimize}\sum_{i=2}^{n} u_i \tag{4.18}$$

subject to the inequalities

$$u_i \geqslant 0, \quad i=2,3,\ldots,n, \tag{4.19}$$

$$\begin{aligned}&\sum_{i=2}^{n}\left[\binom{n}{k}(1-\frac{2k}{n}) - K_k(i)\right] u_i \\ &\quad \leqslant \binom{n}{k}\left[1 - \frac{1}{M^2} + \left(1-\frac{2k}{n}\right)\left(\frac{2^n}{M} - 1 - \frac{2^n-1}{M^2}\right)\right], \quad k=1,2,\ldots,n \end{aligned} \tag{4.20}$$

Note that if $M \geqslant 2^{n-1}-1$, then

$$1 - \frac{1}{M^2} + \left(1-\frac{2k}{n}\right)\left(\frac{2^n}{M} - 1 - \frac{2^n-1}{M^2}\right) \geqslant 0, \quad k=1,2,\ldots,n. \tag{4.21}$$

This implies that $\Lambda_2(n,M)=0$ by choosing the optimal solution $u_2=u_3=\cdots=u_n=0$. Hence, here we only consider the case $M\leqslant 2^{n-1}-1$. Therefore, by (4.16) we obtain that for M odd and $M\leqslant 2^{n-1}-1$,

$$\bar{d}(C)\geqslant \frac{n+1}{2}-\frac{2^{n-1}}{M}+\frac{2^n-n-1}{2M^2}+\frac{1}{2}\Lambda_2(n,M). \tag{4.22}$$

The dual problem of ($\mathscr{B}$) is given as follows:

($\mathscr{B}'$) Choose the real variables $x_1,x_2,\ldots,x_n$ so as to

$$\Lambda_2(n,M)=\text{maximize}\sum_{k=1}^{n}\binom{n}{k}\left[\left(\frac{2k}{n}-1\right)\left(\frac{2^n}{M}-1-\frac{2^n-1}{M^2}\right)-1+\frac{1}{M^2}\right]x_k \tag{4.23}$$

subject to the inequalities

$$x_k\geqslant 0,\quad k=1,2,\ldots,n, \tag{4.24}$$

$$\sum_{k=1}^{n}\left[\binom{n}{k}\left(1-\frac{2k}{n}\right)-K_k(i)\right]x_k\geqslant -1,\quad i=2,3,\ldots,n. \tag{4.25}$$

We have already known that $x_1=x_2=\cdots=x_{n-1}=0, x_n=\frac{1}{2}$ is a feasible solution. Hence,

$$\Lambda_2(n,M)\geqslant\frac{1}{2}\left[\frac{2^n}{M}-1-\frac{2^n-1}{M^2}-1+\frac{1}{M^2}\right]=\frac{2^{n-1}}{M}-1-\frac{2^{n-1}-1}{M^2}.$$

It follows from (4.22) that

Theorem 5. *Let C be a binary (n,M) code. If M is odd and $M\leqslant 2^{n-1}-1$, then*

$$\bar{d}(C)\geqslant\frac{n}{2}-\frac{2^{n-2}}{M}+\frac{2^{n-1}-n}{2M^2}. \tag{4.26}$$

Note that the lower bound in Theorem 5 is only meaningful for M odd and $2^{n-1}/n-1\leqslant M\leqslant 2^{n-1}-1$. It follows from (1.11) that the lower bound in Theorem 5 is tight for $M=2^{n-1}-1$. The following corollary shows that the lower bound in Theorem 5 is also tight for $M=2^{n-2}\pm 1$.

Corollary 5.

$$\beta(n,2^{n-2}-1)=\frac{n-2}{2}-\frac{n-2}{2(2^{n-2}-1)^2}, \tag{4.27}$$

$$\beta(n,2^{n-2}+1)=\frac{n-2}{2}+\frac{2^n-n+2}{2(2^{n-2}+1)^2}. \tag{4.28}$$

Proof. Theorem 5 implies that

$$\beta(n,2^{n-2}-1)\geqslant\frac{n-2}{2}-\frac{n-2}{2(2^{n-2}-1)^2},$$

$$\beta(n,2^{n-2}+1)\geqslant\frac{n-2}{2}+\frac{2^n-n+2}{2(2^{n-2}+1)^2}.$$

Let C_1 and C_2 be the sets obtained by deleting or adding one point from the set $V_{n-2} \times \{00\}$, respectively. It is not hard to check that

$$\bar{d}(C_1) = \frac{n-2}{2} - \frac{n-2}{2(2^{n-2}-1)^2}, \quad \bar{d}(C_2) = \frac{n-2}{2} + \frac{2^n - n + 2}{2(2^{n-2}+1)^2}.$$

Hence

$$\beta(n, 2^{n-2}-1) \leqslant \frac{n-2}{2} - \frac{n-2}{2(2^{n-2}-1)^2},$$
$$\beta(n, 2^{n-2}+1) \leqslant \frac{n-2}{2} + \frac{2^n - n + 2}{2(2^{n-2}+1)^2}$$

and the corollary follows. □

It follows from Lemma 2 and Corollary 5 that

Corollary 6.

$$\beta(n, 2^{n-1} + 2^{n-2} - 1) = \frac{n}{2} - \frac{2^{2n-3} + n}{2(2^{n-1} + 2^{n-2} - 1)^2}, \tag{4.29}$$

$$\beta(n, 2^{n-1} + 2^{n-2} + 1) = \frac{n}{2} - \frac{2^{2n-3} - 2^n + n}{2(2^{n-1} + 2^{n-2} + 1)^2}. \tag{4.30}$$

From Lemma 1 and Theorem 5, we obtain that

Corollary 7. *Let C be a binary (n, M) code, If M is odd and $M \geqslant 2^{n-1} + 1$, then*

$$\bar{d}(C) \geqslant \frac{n}{2} - \frac{2^{n-1}(2^n - M - 1) + n}{2M^2}. \tag{4.31}$$

From (1.12), (4.29) and (4.30), we know that the lower bound in Corollary 7 is tight for $M = 2^{n-1} + 1, 2^{n-1} + 2^{n-2} \pm 1$. Comparing the lower bounds in Theorem 5 and Corollary 7 with the lower bound of Xia and Fu (see Theorem 2), we see that both the lower bounds in Theorem 5 and Corollary 7 are better than the lower bound of Xia and Fu.

5. An upper bound for $\beta(n, M)$

In order to establish our results, we first introduce some notations. For a subset $A \subseteq V_s$ and a vector $b \in V_t$, let

$$A * b = \{(a, b) \mid a \in A\}.$$

Below we use the notation $\mathbf{0}(l)$ to represent the zero vector of length l. For $l = 0$, it represents the empty vector with length 0. Below we give an upper bound for $\beta(n, M)$ by presenting a construction of binary (n, M) codes.

Theorem 6. *For* $1 \leqslant M \leqslant 2^n - 1$, *let the binary expansion of* M *be given by*

$$M = 2^{m_1} + 2^{m_2} + \cdots + 2^{m_k},$$

where $0 \leqslant m_1 < m_2 < \cdots < m_k \leqslant n-1$. *Then*

$$\beta(n,M) \leqslant \frac{1}{M^2}\left[\sum_{i=1}^{k} m_i 2^{2m_i-1} + 2\sum_{1\leqslant i<j\leqslant k} 2^{m_i}(2^{m_j} + m_j 2^{m_j-1})\right] \triangleq \Theta(n,M). \quad (5.1)$$

Proof. Below we construct a binary (n,M) code C with $\bar{d}(C) = \Theta(n,M)$. Let

$$\begin{aligned}
C_k &= V_{m_k} * \mathbf{0}(n-m_k),\\
C_{k-1} &= V_{m_{k-1}} * \mathbf{0}(m_k - m_{k-1}) * 1 * \mathbf{0}(n-1-m_k),\\
C_{k-2} &= V_{m_{k-2}} * \mathbf{0}(m_{k-1} - m_{k-2}) * 1 * \mathbf{0}(m_k - m_{k-1} - 1) * 1 * \mathbf{0}(n-1-m_k),\\
C_{k-3} &= V_{m_{k-3}} * \mathbf{0}(m_{k-2} - m_{k-3}) * 1 * \mathbf{0}(m_{k-1} - m_{k-2} - 1) * 1 *\\
&\qquad \mathbf{0}(m_k - m_{k-1} - 1) * 1 * \mathbf{0}(n-1-m_k),\\
&\quad\ \vdots\\
C_1 &= V_{m_1} * \mathbf{0}(m_2 - m_1) * 1 * \mathbf{0}(m_3 - m_2 - 1) * 1 * \mathbf{0}(m_4 - m_3 - 1) * \cdots *\\
&\qquad 1 * \mathbf{0}(m_k - m_{k-1} - 1) * 1 * \mathbf{0}(n-1-m_k).
\end{aligned}$$

If $m_1 = 0$, the set V_0 represents the set which only contains the empty vector. It is easy to see that the sets $C_1, C_2, \ldots, C_k$ are disjoint of each other and $|C_i| = 2^{m_i}$, $i = 1,2,\ldots,k$. Let $C = C_1 \cup C_2 \cup \cdots \cup C_k$, then C is a binary (n,M) code. Next, we show that $\bar{d}(C) = \Theta(n,M)$.

$$\sum_{a,b\in C} d_{\mathrm{H}}(a,b) = \sum_{i=1}^{k}\sum_{a,b\in C_i} d_{\mathrm{H}}(a,b) + 2\sum_{1\leqslant i<j\leqslant k}\sum_{\substack{a\in C_i\\ b\in C_j}} d_{\mathrm{H}}(a,b). \quad (5.2)$$

From the construction of C_i, we have

$$\begin{aligned}
\sum_{a,b\in C_i} d_{\mathrm{H}}(a,b) &= \sum_{x,y\in V_{m_i}} d_{\mathrm{H}}(x,y) = 2^{m_i}\sum_{l=0}^{m_i} l\binom{m_i}{l}\\
&= 2^{m_i}\cdot m_i \cdot 2^{m_i-1} = m_i \cdot 2^{2m_i-1}. \quad (5.3)
\end{aligned}$$

From the constructions of C_i and C_j, we have that for $1 \leqslant i < j \leqslant k$ and any $a \in C_i$,

$$\begin{aligned}
\sum_{b\in C_j} d_{\mathrm{H}}(a,b) &= \sum_{b\in C_j} w_{\mathrm{H}}(a+b) = \sum_{x\in V_{m_j}} [1 + w_{\mathrm{H}}(x)] = 2^{m_j} + \sum_{x\in V_{m_j}} w_{\mathrm{H}}(x)\\
&= 2^{m_j} + \sum_{l=0}^{m_j} l\binom{m_j}{l} = 2^{m_j} + m_j \cdot 2^{m_j-1}. \quad (5.4)
\end{aligned}$$

Therefore,

$$\sum_{a\in C_i}\sum_{b\in C_j} d_{\mathrm{H}}(a,b) = 2^{m_i}[2^{m_j} + m_j 2^{m_j-1}]. \quad (5.5)$$

From (5.2), (5.3) and (5.5), we obtain that $\bar{d}(C) = \Theta(n,M)$. This implies that $\beta(n,M) \leqslant \bar{d}(C) = \Theta(n,M)$. □

We can check that $\beta(n,M) = \Theta(n,M)$ for many specific parameters n and M. This led us to conjecture that $\beta(n,M) = \Theta(n,M)$ for all n and $1 \leqslant M \leqslant 2^n$ in the early version of this paper. A referee and Kündgen pointed out that the conjecture is not true in general. The referee gave a counterexample as follows. Let $M = 2^k$, where $k \geqslant 4$ and $2^k \leqslant n$. Then $\Theta(n, 2^k) = k/2$. Now take

$$C = \{e_1, e_2, \ldots, e_M\}.$$

Then

$$\bar{d}(C) = 2(M-1)/M < 2 \leqslant k/2.$$

This implies that the conjecture is not true in general. Furthermore, it was mentioned by Kündgen (see also [6]) that when $M < n$, then the optimal configuration (for $M > 8$ unique) for achieving $\beta(n,M)$ is to take the points $\mathbf{0}, e_1, e_2, \ldots, e_{M-1}$. Hence, for $M < n$,

$$\beta(n,M) = 2(M-1)^2/M^2.$$

By Lemma 2, we know that for $M < n$,

$$\beta(n, 2^n - M) = \frac{n}{2} - \frac{nM^2 - 4(M-1)^2}{2(2^n - M)^2}.$$

Acknowledgements

The authors wish to thank the anonymous referees and Dr. A. Kündgen for very useful suggestions that improved the presentation of the results in this paper.

References

[1] R. Ahlswede, I. Althöfer, The asymptotic behaviour of diameters in the average, J. Combin. Theory Ser. B 61 (1994) 167–177.
[2] R. Ahlswede, G. Katona, Contributions to the geometry of Hamming spaces, Discrete Math. 17 (1977) 1–22.
[3] R. Ahlswede, Z. Zhang, Coding for write-efficient memory, Information and Computation 83 (1994) 80–97.
[4] I. Althöfer, T. Sillke, An "average distance" inequality for large subsets of the cube, J. Combin. Theory Ser. B 56 (1992) 296–301.
[5] M.R. Best, A.E. Brouwer, F.J. MacWilliams, A.W. Odlyzko, N.J.A. Sloane, Bounds for binary codes of length less than 25, IEEE Trans. Inform. Theory IT-24 (1981) 81–93.
[6] P. Delsarte, Bounds for unrestricted codes, by linear programming, Philips Res. Rep. 27 (1972) 272–289.
[7] F.-W. Fu, T. Kløve, S.-Y. Shen, On the Hamming distance between two i.i.d. random n-tuples over a finite set, IEEE Trans. Inform. Theory IT-45 (1999) 803–807.
[8] F.-W. Fu, S.-Y. Shen, On the expectation and variance of Hamming distance between two binary i.i.d. random vectors, Acta Math. Appl. Sinica 13 (1997) 243–250.

[9] A. Kündgen, Covering cliques with spanning bicliques, J. Graph Theory 27 (1998) 223–227.
[10] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, New York, 1985.
[11] S.-T. Xia, F.-W. Fu, On the average Hamming distance for binary codes, Discrete Appl. Math. 89 (1998) 269–276.